

REPORT DOCUMENTATION PAGE			Form Approved OMB NO. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comment regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.				
1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE 23 FEB 99		3. REPORT TYPE AND DATES COVERED Final Progress 1 MAR 95 - 31 DEC 99
4. TITLE AND SUBTITLE Interacting Sites in Novel Polymeric Proteins			5. FUNDING NUMBERS DAAH04-95-1-0129	
6. AUTHOR(S) Steven T. Case				
7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(ES) Department of Biochemistry University of Mississippi Medical Center 2500 North State Street Jackson, MS 39216			8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211			10. SPONSORING / MONITORING AGENCY REPORT NUMBER ARO 33106.5-LS	
11. SUPPLEMENTARY NOTES The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.				
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution unlimited.			12 b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words) The object of this project was the folding and association of Cys-rich recombinant and native silk proteins from aquatic larvae of the midge, <i>Chironomus tentans</i> . Specific aims include. A. Learn which Cys residues are critical to the pathway for intramolecular disulfide bond formation in rCAS, a recombinant repeat from silk protein, sp1a. B. Determine if intramolecular disulfide bonds stabilize the higher-order structure of rCAS C. Determine the conditions and which Cys participate in formation of intermolecular disulfide bonds. D. Learn how and where other native silk proteins interact with sp1s. In addition, we also chose to do the following. E. Determine if N-linked glycosylation sites in ssp160 are evolutionarily conserved. F. Acquire consensus repeat sequences for Cys-containing motifs in sp185/sp220 homologs and sp195. The final results of this project indicate: 1) Aquatic silk proteins can be expressed and purified from bacteria in large quantities. 2) However, fibrous proteins may prove difficult to analyze: CD and FTIR basis spectra lack fibrous protein standards and proteins with Cys may refold into multiple conformers. 3) The abundance and conservation of glycosylation sites on some aquatic silk proteins suggest such post-translational modification merits study before biotechnological applications are considered. 4) Aquatic silk proteins contain more Cys motifs than previously thought.				
14. SUBJECT TERMS silk proteins, cysteine, disulfide bonds, intraprotein interactions, protein-protein interactions			15. NUMBER IF PAGES 7	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OR REPORT UNCLASSIFIED	18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED	20. LIMITATION OF ABSTRACT UL	

Steven T. Case

FINAL PROGRESS REPORT INTERACTING SITES IN NOVEL POLYMERIC PROTEINS

1. FORWARD

The primary sequence of a protein leads to its higher-order (secondary, tertiary and quaternary) structure; sites of intraprotein and protein-protein interactions are contained therein. Higher-order structures and properties of a native protein are acquired through protein folding pathways that are best understood for small globular proteins¹⁻⁴. Fibrous proteins comprise essential building blocks for extracellular structures⁵. The chemical and physical properties of several (such as collagen, keratin and silk) are known and provide the basis for novel protein-based biomolecular materials⁶. However, in contrast to globular proteins, little is known about the structure and folding pathways of fibrous proteins.

Aquatic larvae of the midge, *Chironomus tentans*, spin silk. In contrast to spiders and silkworms, midge silk contains 7% cysteine (Cys)⁷. Six midge silk proteins are composed of 50-130 copies of tandemly repeated sequences with two or four invariant Cys residues⁸. The evolutionary conservation of these residues implies functional significance. This notion is supported by the fact Cys residues in a bacterially expressed core repeat from silk protein spIa can form *intramolecular* disulfide bonds *in vitro*⁹ and native silk proteins form *intermolecular* disulfide bonds *in vivo*¹⁰. These residues are therefore sites of intraprotein and protein-protein interactions and provide a unique opportunity to study the folding pathway and associative properties of fibrous proteins.

Knowledge gained about the structure and interactive sites of aquatic midge silk proteins will further impact the potential biotechnological application of these proteins as novel biomolecular materials.

2. TABLE OF CONTENTS

Forward.....	1
Table of Contents.....	1
List of Appendixes and Tables.....	1
Statement of Problem Studied.....	2
Summary of Most Important Results.....	2
Publications and Technical Reports.....	4
Participating Scientific Personnel.....	5
Report of Inventions.....	5
Bibliography.....	5
Appendixes.....	6

3. LIST OF APPENDIXES AND TABLES

- Figure 1: Disulfide Linkages in rCAS, a Recombinant Aquatic Silk Protein
- Figure 2: Dissociation of Aquatic Silk Protein Complexes
- Figure 3: Cysteine-Containing Motifs in Aquatic Silk Proteins

4. STATEMENT OF PROBLEM STUDIED

The objective of this project was to study the structure, folding and association of midge silk proteins. We initially focused on Cys residues that provide sites of intraprotein and protein-protein interactions in rCAS, a recombinant protein modeled after core repeats in midge silk protein, sp1a⁹. We also examined which silk proteins associate with spIs. Experiments outlined in this proposal were designed to answer the following questions.

- A. Which Cys residues are critical to the pathway for *intramolecular* disulfide bond formation in rCAS? [*Answer: all can participate, order of priority unknown.*]
- B. Are *intramolecular* disulfide bonds required to stabilize the higher-order structure of rCAS? [*Answer: not testable since unable to ascertain the higher order structure.*]
- C. What conditions promote, and which Cys participate in, formation of *intermolecular* disulfide bonds? [*Answer: some pairwise associations preferred.*]
- D. How and where do other silk proteins interact with spIs? [*Answer: one or few molecules of sp140 and sp185 form intermolecular disulfide bonds with spIs; limited number implies sites are Cys within terminal domains rather than internal arrays of repeats.*]

Unanticipated problems were encountered with rCAS, namely conflicting structural predictions and the inability to refold rCAS into a single isomer. However, two new aspects of this project emerged by studying additional silk proteins along the way. They enabled us to answer the following questions.

- E. What evidence is there that glycosylation of aquatic silk proteins is important? [*Answer: number and location of N-linked glycosylation motifs in ssp160 conserved among two species in spite of genetic polymorphism.*]
- F. Do sp185/sp220 homologs and sp195 contain repeats with conserved Cys? [*Answer: sp185/sp220 are non-repetitive but contain 72 blocks of 20-28 residues with a novel Cys-containing motif whereas sp195 is composed of tandem arrays of 25 residue repeats that include yet another novel Cys-containing motif.*]

5. SUMMARY OF MOST IMPORTANT RESULTS

A. Recombinant Silk Proteins

We successfully produced and purified recombinant rCAS derivatives with all possible combinations of zero to four Cys to Ala substitutions (acquired by *in vitro* mutagenesis of the synthetic rCAS-encoding gene).

CD and FITR spectra were obtained. All proteins had significant secondary structure; however, these methods significantly contradicted each other's prediction of the percentage of helix, turn and sheet. Since basis spectra data for both methods were derived from globular proteins, we concluded neither is suitable until characterized fibrous protein standards exist.

Solvent and thermal denaturation studies brought more conflicts; rather than cooperative denaturation curves, linear curves we obtained suggesting all proteins were random coils. Identical results were obtained using proteins with and without Cys, negating the use of this method to study how intramolecular disulfide bonds stabilized secondary structure.

One positive result was obtained: upon reduction and re-oxidation, all two-Cys-containing

derivatives quantitatively reformed one intramolecular disulfide bond. This indicates either unparalleled flexibility or artifactual refolding of rCAS derivatives. At higher concentrations, some showed preference for formation intermolecular disulfide bonds (Fig. 1). When three or more Cys were present, HPLC indicated refolding resulted in multiple conformations. This negated all chances to perform higher order structural studies such as NMR and crystallography.

B. Native Silk Protein Complexes

Progress was made studying native silk protein complexes (listed below in section 6 as publication P3). We determined the approximate mass (2400 kDa) and first qualitative description of the composition of a fibrous protein quaternary structure. This complex contains spI, sp185, sp140 covalently linked by disulfide bonds and possibly sp40, sp17 and sp12 linked non-covalently (Fig. 2). The 130 nm diameter lattice-like complexes form oligomers that can associate into multi-stranded beaded fibers in a concentration-dependent manner. Disulfide bond reduction dissociates sp185 and sp140 from spIs, supporting the notion that Cys conserved in these proteins are also the site of protein-protein interactions.

Analytical ultracentrifugation studies were not feasible due to insolubility of silk protein complexes in UV-transparent buffers required to profile the proteins at low wavelengths (aquatic silk proteins contain relatively few aromatic amino acids)

C. Conservation of Glycosylation Motifs in a Silk Protein

We suspect the "beads" on silk fibers assembled from high molecular mass complexes *in vitro* are carbohydrate chains. Glycosylation of aquatic silk proteins has been reported but largely ignored. We therefore pursued one silk protein whose glycosylation may be extreme.

ssp160 is the special lobe-specific 160-kDa silk protein whose synthesis is limited to four cells that surround the salivary duct where fibers form *in vivo*. Lectin binding and affinity chromatography indicate this silk protein contains both N- and O-linked sugar¹¹. ssp160 cDNA was first isolated from *Chironomus thummi*¹¹, a species for which little molecular biological data exist for silk proteins; this protein has never been observed in *C. tentans*, the species with the largest database. We chose to bridge this gap and, in doing so, answered two classical questions that were posed in this biological system over 30 years ago.

C. thummi ssp160 cDNA was used to isolate the corresponding gene (publication P2) and homologous cDNA and gene from *Chironomus pallidivittatus* (publications P5 and P6), the closest relative to *C. tentans*. Their comparison revealed evolutionary conservation of 12-13 copies of an N-linked glycosylation motif sequestered in two regions of ssp160 where codon deletions are frequent (publication P5). These deletions appear to result from slipped-strand mispairing among arrays of [ACA] repeats. The *C. pallidivittatus* sp160 gene mapped to Balbiani ring 4 (BR4) on polytene chromosome IV, the last BR without an identifiable gene. This result fulfilled Beermann's prediction that all BRs contain a gene encoding a major tissue-specific (silk) protein.

Comparison of an ssp160 gene from *C. thummi* (publication P2) and two alleles from *C. pallidivittatus* (publication P6) revealed differences in intron sizes and downstream flanks. All are attributable to slipped-strand mispairing among 5-bp repeats contained therein

(publication P6). Large-scale but similar dynamics may have resulted in gene deletion: probes for the *C. pallidivittatus* gene and flanks were used to demonstrate that the apparent absence of BR4 on *C. tentans* polytene chromosomes is due to the fact that the ssp160 gene was deleted (publication P7). The upstream deletion breakpoint has been mapped with nucleotide precision; the downstream breakpoint is less clear due to repeats.

The above study increased our need for an efficient DNA sequencing strategy. Thus we invested time in combining a new plasmid vector for transposon-facilitated DNA sequencing with long and accurate PCR mapping of transposon insertion sites (publication P4).

D. Other Novel Cys-Containing Motifs

We completed the full-length cDNA sequence of *C. pallidivittatus* sp185 and *C. thummi* sp220 (publication P1) and compared them to *C. tentans* sp185. These silk proteins contain 72 blocks of 20-28 residues, 61 % of which contain the novel Cys-containing motif: (X₅₋₈)-Cys-(X₅)-(Trp/Phe/Tyr)-(X₄)-Cys-X-Cys-X-Cys.

Only 120 nucleotides of cDNA have ever been reported¹² for sp195 but they encode two Cys residues. We therefore decided to attempt to complete the 6-kb cDNA sequence encoding this protein. Tandem arrays of 75-bp protein-coding repeats within this cDNA render it unstable when propagated in bacterial plasmids. Nonetheless we have acquired over 4 kb of cDNA sequence and found blocks of nearly perfect 25-residue repeats that contain yet another novel Cys-containing motif: DIPANKKWNENTC(C/S)LECKT(E/V)KPKP(D/Q) (publication P8).

E. Summary

Results obtained from this project have established the following. 1) Aquatic silk proteins can be expressed and purified from bacteria in large quantities, however, fibrous proteins may be unexpectedly difficult to analyze due to lack of basis spectra and those with Cys may be susceptible to formation of multiple conformers during refolding. 2) Aquatic silk proteins contain more Cys motifs than previously thought (Fig. 3). 3) The abundance and conservation of glycosylation sites on some aquatic silk proteins suggest such post-translational modification merits study before biotechnological applications are considered. The same may apply to silks in general; spider silk is glycosylated too.

6. PUBLICATIONS AND TECHNICAL REPORTS

P1. Case, S.T. Cox, C., Bell, W.C., Hoffman, R.T., Martin, J. and Hamilton, R. Extraordinary conservation of cysteines among homologous *Chironomus* silk proteins sp185 and sp220. J. Mol. Evol. 44: 452-462 (1997).

P2. Berezikov, E., Blinov, A.G., Scherbik, S., Cox, C.K. and Case, S.T. Structure and polymorphism of the *Chironomus thummi* gene encoding special lobe-specific silk protein, ssp160. Gene 223:347-354 (1998).

Steven T. Case

- P3. Case, S.T. and Thorton, J.R. High molecular mass complexes of aquatic silk proteins. *Int. J. Biol. Macromol.* (in press, 1999).
- P4. Cox, C.K., Anido, A.E. and Case, S.T. Efficient transposon-facilitated DNA sequencing of large target DNAs. (submitted)
- P5. Case, S.T., Cox, C.K. and Anido, A.I. The last Balbiani ring: BR4 in *Chironomus pallidivittatus* encodes ssp160, a special lobe-specific 160-kDa silk protein. (submitted)
- P6. Cox, C.K., Anido, A.E. and Case, S.T. Polymorphic alleles of the *Chironomus pallidivittatus* gene encoding ssp160. (in preparation)
- P7. Anido, A.E., Cox, C.K. and Case, S.T. A molecular explanation for the fate of BR4 in *Chironomus tentans*. (in preparation)
- P8. Case, S.T., Goel, A., Cox, C.K., Huff, E. and Donhardt, A. Novel cysteine-containing motif in *Chironomus tentans* silk protein, sp195. (in preparation)

Anido, A.E. (1996) "Use of a Novel Plasmid for Transposon-Mediated DNA Sequencing", Honor's Thesis, Millsaps College, Jackson, MS .

Donhardt, A.M. (1998) "cDNA for a 195-kDa Silk Protein from Larval Salivary Glands of *Chironomus tentans*", Honors Thesis, Mississippi College, Clinton, MS.

7. PARTICIPATING SCIENTIFIC PERSONNEL

Jennifer R. Thornton,
Sudha Govindichar
Carol K. Cox
Undergraduate AASERT Awardees
Amy Donhardt,
Aimee E. Anido,
Elizabeth Huff,
Andrew O'Dell,

8. REPORT OF INVENTIONS - None

9. BIBLIOGRAPHY

1. Afinsen, C.B. (1973) *Science* 181:223-230
2. Rothwarf, D.M., and Sheraga, H.A., (1993) *Biochemistry* 32:2671-2679
3. Creighton, T.E. and Goldenberg, D.P. (1984) *J. Mol. Biol.* 179:497-526
4. Weismann, J.S. and Kim, P.S. (1991) *Science* 253:1386-1393

Steven T. Case

5. *Fibrous Protein Structure*, (J.M. Squire and P.J. Vibert, eds.) Academic Press, NY, 1987, pp.1- 535
6. *Biomolecular Materials*, Materials Research Society Symposium Proceedings, Pittsburgh, PA, 1993, 232:1-288
7. Case, S.T., Powers, J.M., Hamilton, R., Burton, M.J. (1994) in *Silk Polymers: Materials Science and Biotechnology*, (D. Kaplan et al., eds.), American Chemical Society Symposium 544, Washington, pp. 91-97
8. Case, S.T. and Wieslander, L.W. (1992) Res. Prob. Cell Diff. 19:187-226
9. Smith, S.V., Correia, J.J. and Case, S.T. Prot. Sci. (1995) 4:945-954.
10. Case, S.T. and Thorton, J.R. High molecular mass complexes of aquatic silk proteins. Int. J. Biol. Macromol. (in press, 1999).
11. Hoffman, R.T., Schmidt, E.R. and Case, S.T. (1996) J. Biol. Chem. 271:9809-9815.
12. Dreesen, T.D., Bower, J.R. and Case, S.T. (1985) J. Biol. Chem. 60:11824-11830.

10. APPENDIXES -

Attached are 3 figures.

Steven T. Case

Figure 1. Disulfide Linkages in rCAS.

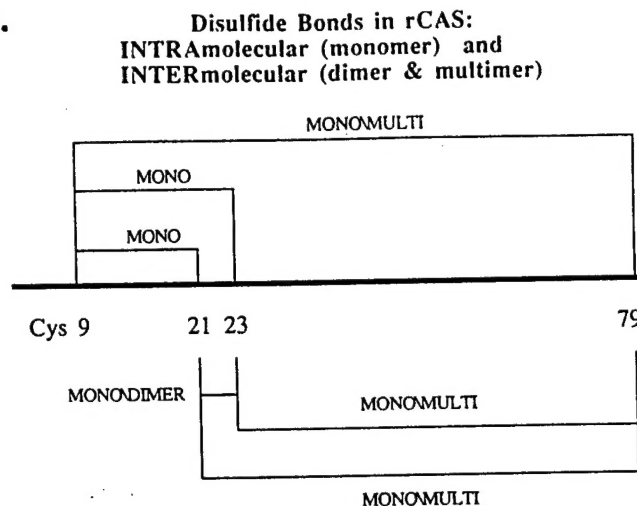


Figure 2. Dissociation of aquatic silk proteins.

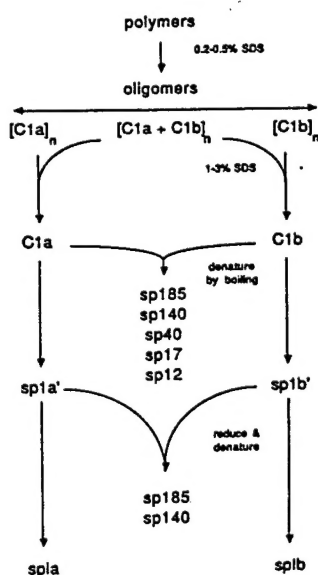


Figure 3. Cys-containing motifs in aquatic silk proteins.

<u>Silk Protein</u>	<u>Motif</u>	<u>Copies</u>
sp1a, sp1b, sp1c, sp1d	C-X ₁₁ -C-X ₆ -F-X ₄ -C-X-C-X ₂	60-90
sp185 and sp220	X ₅ -8-C-X ₅ -(W/F/Y)-X ₄ -C-X-C-X-C	72
sp195	X ₁₂ -C-(C/S)-L-E-C-X ₈	~80